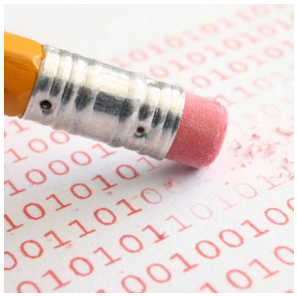


Machine unlearning

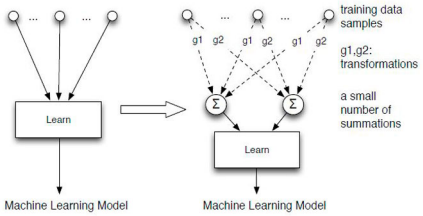
Publié le lundi 28 mars 2016
Voir en ligne : https://www.france-science.org/Machine-unlearning.html

Les algorithmes de machine learning sont devenus très populaires grâce à leurs capacités de prédiction dans tous les domaines (risques naturels, loisirs, santé, …) en exploitant les différentes relations existantes entre les données. Cependant, ce type d'approche utilise de manière intensive les données des utilisateurs, parfois à leur insu. Il est donc intéressant de prévoir des solutions pour contrôler la puissance de ces algorithmes, assurer des propriétés liées à la sécurité et offrir une certaine confidentialité des données.

Ces nouvelles propriétés sont notamment préconisées par le « droit à l'oubli » qui a été introduit comme un droit fondamental en Europe et qui a fait l'objet d'une proposition de réglementation européenne en 2014. Cette problématique est de plus en plus présente et Google en 2014 a dû supprimer plus de 170 000 liens pour satisfaire aux exigences de respect de la vie privée [1].

Yinzhi Cao et Junfeng Yang, chercheurs respectivement à Lehigh University et Columbia University, ont obtenu de la NSF une bourse de 1,2 millions de dollars pour développer une machine dite « unlearning », c'est à dire une machine capable de désapprendre ! [2]

Leur méthode plus rapide et efficace que les logiciels existants permet à un système d'apprentissage automatique d'oublier voire de supprimer des données en s'appuyant sur un modèle évoluant progressivement.



Leur approche consiste à introduire des données agrégées dans le système d'apprentissage par une couche logicielle intermédiaire qui permet de ne pas donner accès aux relations ou dépendances entre les données brutes et donc d'avoir une sorte de mécanisme d'oubli de ces données initiales.

Quatre systèmes existants ont été testés avec succès : Lenskit, un système de recommandation open-source, Zozzle, un détecteur de malware, un filtre spam utilisé sur les réseaux sociaux, ainsi que PJScan, un autre détecteur de malware.
La phase suivante du projet va consister à spécifier et développer des outils pour vérifier l'efficacité de la méthode sur tout système.

Toute donnée et plus précisément vos données ont une valeur. Il est important que la communauté scientifique offre les outils nécessaires pour redonner aux utilisateurs le contrôle et la liberté de choix sur l'utilisation de leurs données.

Rédacteurs :
- Hervé Martin, Attaché pour la Science et la Technologie, attache-ntics@ambascience-usa.org
- Marie Letoret, Attachée adjointe pour la Science et la Technologie, deputy-ntics@ambascience-usa.org

Notes

[1] https://fr.wikipedia.org/wiki/Droit_%C3%A0_l%27oubli

[2] http://www1.lehigh.edu/news/new-technique-wipes-out-unwanted-data